



# Analisis Sentimen Pengunjung Pulau Komodo dan Pulau Rinca di Website Tripadvisor Berbasis CRISP-DM

Yerik Afrianto Singgalen\*

Fakultas Ilmu Administrasi Bisnis dan Ilmu Komunikasi, Program Studi Pariwisata, Universitas Katolik Indonesia Atma Jaya, Jakarta

Jln. Jend. Sudirman No.51, RW.4, Karet Semanggi, Kecamatan Setiabudi, Kota Jakarta Selatan, Daerah Khusus Ibukota Jakarta, Indonesia

Email: <sup>1,\*</sup>yerik.afrianto@atmajaya.ac.id

Email Penulis Korespondensi: yerik.afrianto@atmajaya.ac.id

Submitted: 22/01/2023; Accepted: 30/01/2023; Published: 31/01/2023

**Abstrak**—Analisis sentimen diperlukan untuk mengidentifikasi preferensi wisatawan terhadap produk dan layanan di suatu destinasi wisata. Oleh sebab itu penelitian ini menggunakan metode Cross-Industry Standard Procedure for Data Mining (CRISP-DM) untuk mengklasifikasi data ulasan pengunjung Pulau Komodo dan Pulau Rinca yang diperoleh dari website Tripadvisor berdasarkan sentimen negatif dan positif, kemudian merekomendasikan pengembangan sistem informasi yang dapat mengoptimalkan produk dan layanan destinasi wisata di Pulau Komodo dan Pulau Rinca. Adapun, algoritma yang digunakan ialah k-Nearest Neighbor (k-NN), Naïve Bayes Classifier (NBC), Support Vector Machine (SVM), dan Decision Tree (DT). Berdasarkan hasil klasifikasi data menggunakan metode pembobotan kata atau Term Frequency Inverse Document Frequency (TF-IDF) melalui operator ekstrak sentimen dapat diketahui lima kata teratas yang paling sering muncul dalam ulasan wisatawan terhadap Pulau Komodo ialah sebagai berikut: komodo (1894), dragons (1596), island (1492), tour (840), boat (774). Adapun, lima kata teratas yang paling sering muncul dalam ulasan wisatawan terhadap Pulau Rinca ialah sebagai berikut: komodo (1042), dragons (962), island (882), rinca (606), boat (372). Disisi lain, hasil klasifikasi 564 data ulasan wisatawan Pulau Komodo dan 364 data ulasan wisatawan Pulau Rinca menggunakan algoritma k-NN, NBC, SVM, dan DT berbasis CRISP-DM, menunjukkan bahwa Support Vector Machine (SVM) merupakan algoritma dengan performa terbaik dimana nilai akurasi 99,69%, presisi 100%, recall 99,39%, f-measure 99,69%, Area Under Curve (AUC) 100% dan t-Test 0,958. Selanjutnya, hasil pengolahan data ulasan wisatawan terhadap produk dan layanan Pulau Rinca, menunjukkan bahwa algoritma Support Vector Machine (SVM) memiliki performa terbaik dengan nilai akurasi 100%, presisi 100%, recall 100%, f-measure 100%, Area Under Curve (AUC) 100% dan t-Test 0,964. Setelah membandingkan performa SVM sebelum dan setelah menggunakan Synthetic Minority Oversampling Technique (SMOTE), dapat diketahui bahwa performa algoritma menjadi lebih optimal apabila menggunakan operator SMOTE. Dengan demikian, SVM merupakan algoritma yang relevan digunakan sebagai model analisis sentimen wisatawan di Pulau Komodo dan Pulau Rinca berdasarkan CRISP-DM.

**Kata Kunci:** Naïve Bayes; Support Vector Machine; Decision Tree; k-NN; Analisis Sentimen; CRISP-DM

**Abstract**—A sentiment analysis is needed to identify tourist preferences for products and services in a tourist destination. Therefore, this study uses the Cross-Industry Standard Procedure for Data Mining (CRISP-DM) method to classify visitor review data for Komodo Island and Rinca Island obtained from the Tripadvisor website based on negative and positive sentiments, then recommends the development of an information system that can optimize the products and services of tourist destinations on Komodo Island and Rinca Island. Meanwhile, the algorithms used are k-Nearest Neighbor (k-NN), Naïve Bayes Classifier (NBC), Support Vector Machine (SVM), and Decision Tree (DT). Based on the results of data classification using the word weighting method or Term Frequency Inverse Document Frequency (TF-IDF) through the sentiment extract operator, it can be known that the five words that most often appear in tourist reviews of Komodo Island are as follows: Komodo dragons (1894), dragons (1596), island (1492), tour (840), boat (774). Meanwhile, the five words that most often appear in tourist reviews of Rinca Island are as follows: komodo dragon (1042), dragons (962), island (882), rinca (606), and boat (372). In addition, the results of the classification of 564 Komodo Island tourist review data and 364 Rinca Island tourist review data using CRISP-DM-based k-NN, NBC, SVM, and DT algorithms, show that the Support Vector Machine (SVM) is the best-performing algorithm where the accuracy value is 99.69%, precision 100%, recall 99.39%, f-measure 99.69%, Area Under Curve (AUC) 100% and t-Test 0.958. Meanwhile, the results of processing tourist review data on Rinca Island products and services show that the Support Vector Machine (SVM) algorithm has the best performance with 100% accuracy, 100% precision, 100% recall, 100% f-measure, 100% Area Under Curve (AUC) and 0.964 t-Test. After comparing the performance of SVM before and after using the Synthetic Minority Oversampling Technique (SMOTE), it can be seen that the implementation of the algorithm becomes more optimal when using the SMOTE operator. Thus, SVM is a relevant algorithm used as a model for analyzing tourist sentiment on Komodo Island and Rinca Island based on CRISP-DM.

**Keywords:** Naïve Bayes; Support Vector Machine; Decision Tree; k-NN; Sentiment Analysis; CRISP-DM

## 1. PENDAHULUAN

Salah satu destinasi wisata prioritas nasional di Indonesia yang termasuk tujuh keajaiban dunia ialah Pulau Komodo. Anju et al. menunjukkan bahwa destinasi wisata Pulau Komodo memiliki daya tarik kunjungan wisatawan mancanegara yang meningkatkan perekonomian Labuan Bajo [1]. Disisi lain, Dima dan Ketmoen menunjukkan bahwa total penerimaan daerah dari Taman Nasional Komodo dan sektor pariwisata tahun 2020 mencapai 1.431.100.000 meskipun dilanda covid-19 [2]. Tulipa et al. menunjukkan bahwa perkembangan pariwisata di Pulau Komodo dan Pulau Rinca juga memengaruhi perkembangan sektor pariwisata Manggarai, Nusa Tenggara Barat [3]. Destinasi wisata Pulau Komodo dan Pulau Rinca ramai dikunjungi wisatawan nusantara



dan mancanegara sehingga memantik pertumbuhan ekonomi dan sosial masyarakat lokal. Adapun, Arifiyanti et al. menunjukkan bahwa perkembangan teknologi informasi berbasis website memungkinkan wisatawan sebagai pengguna sistem untuk mengulas produk dan layanan selama berada di destinasi wisata, menanggapi hal tersebut pihak pengelola destinasi akan berupaya menjaga citra baik destinasi dengan mempertahankan atau meningkatkan kualitas produk dan layanan di destinasi wisata [4]. Afidah et al. menunjukkan bahwa salah satu pendekatan yang dapat digunakan untuk mengidentifikasi, mengklasifikasi dan menganalisis kepuasan wisatawan di website Tripadvisor ialah pendekatan analisis sentimen melalui proses data mining [5]. Hal ini menunjukkan bahwa pendekatan Knowledge Discovery and Data-Mining dapat digunakan untuk mengklasifikasi sentimen wisatawan. Dengan demikian, penelitian ini bertujuan untuk menganalisis sentimen wisatawan menggunakan pendekatan Knowledge Discovery in Databases (KDD) dan Cross-Industry Standard Process for Data Mining (CRISP-DM) dalam klasifikasi data ulasan wisatawan di website Tripadvisor pada destinasi wisata Pulau Komodo dan Pulau Rinca, Manggarai, Nusa Tenggara Barat, Indonesia.

Pendekatan data mining dan knowledge discovery menekankan pada eksplorasi data serta pengolahan data menjadi informasi yang bermanfaat untuk inovasi bisnis dan teknologi. Rohini dan Marraynal berpendapat bahwa teknik data mining umumnya digunakan pada dataset dengan jumlah data yang sangat banyak dan tidak dapat dilakukan secara manual [6]. Disisi lain, Rahman et al. menunjukkan bahwa model KDD juga mengalami perkembangan menjadi CRISP-DM karena paling banyak digunakan untuk mengembangkan organisasi profit [7]. Disisi lain, Zauardi dan Suprayitno menunjukkan bahwa model KDD memiliki alur yang berbeda dengan CRISP-DM, dimana KDD memiliki sembilan tahapan sebagai berikut: selection and addition; pre-processing data; transformation; data mining; evaluation and interpretation. Sedangkan CRISP-DM memiliki alur sebagai berikut : business understanding; data understanding; data preparation; modeling; evaluation; deployment. Hal ini menunjukkan bahwa KDD dapat digunakan pada organisasi nirlaba sedangkan CRISP-DM lebih relevan digunakan untuk kepentingan organisasi profit. Dengan demikian, penelitian ini akan menggunakan pendekatan CRISP-DM dengan mempertimbangkan relevansi konteks pembahasan di sektor pariwisata untuk pengembangan sosial dan ekonomi lokal.

Sentimen wisatawan berperan penting dalam meningkatkan citra destinasi wisata. Ginantra et al. menunjukkan bahwa analisis sentimen destinasi di sektor pariwisata perlu ditingkatkan untuk memberikan rekomendasi kepada pengelola destinasi maupun akomodasi di lokasi wisata agar mengoptimalkan kualitas produk dan layanan bagi wisatawan [8]. Disisi lain, Wahyu et al. menunjukkan bahwa perkembangan teknologi informasi memudahkan akses terhadap data sentimen wisatawan melalui platform digital berbasis website yang digunakan oleh pengusaha pariwisata sebagai strategi pemasaran [9]. Adapun, Arsa et al. berpendapat bahwa analisis sentimen merupakan strategi yang relevan digunakan menanggapi perubahan permintaan wisatawan terhadap karakteristik produk dan layanan di bisnis pariwisata, terutama sebelum dan setelah pandemi Covid-19 [10]. Hal ini menunjukkan bahwa analisis sentimen merupakan pendekatan baru yang perlu dikembangkan sebagai bagian dari strategi pemasaran digital bisnis pariwisata. Dengan demikian, penelitian ini akan melakukan analisis sentimen di sektor pariwisata berdasarkan studi kasus di Pulau Komodo dan Pulau Rinca menggunakan dataset dari website Tripadvisor berbasis CRISP-DM.

Penelitian tentang analisis sentimen di Indonesia perlu dilakukan untuk mengidentifikasi dan memetakan preferensi wisatawan terhadap produk dan layanan wisata berdasarkan konteks destinasi di masing-masing daerah, penelitian ini merupakan bagian dari roadmap penelitian untuk kajian analisis sentimen di destinasi wisata prioritas [11]–[16]. Secara spesifik, analisis sentimen dalam penelitian ini fokus mengkaji sentimen wisatawan terhadap produk dan layanan wisata di Pulau Komodo dan Pulau Rinca sebagai destinasi wisata prioritas di Indonesia. Adapun, sumber data teks diambil dari Website Tripadvisor, sedangkan algoritma yang digunakan dalam klasifikasi data teks terdiri dari algoritma k-Nearest Neighbor (k-NN), Naïve Bayes Classifier (NBC), Support Vector Machine (SVM), dan Decision Tree (DT). Mahardika et al. menunjukkan bahwa hasil klasifikasi sentimen wisatawan terhadap produk dan layanan dapat dikembangkan menjadi dashboard berbasis website untuk merekomendasikan program pengembangan destinasi yang kontekstual terhadap pemangku kepentingan yang bertanggungjawab di bidang pariwisata daerah [17]. Selain itu, Somantri dan Dairoh menunjukkan bahwa hasil sentimen wisatawan merupakan perspektif konsumen yang perlu dianalisis dan diseimbangkan dengan perspektif pengelola destinasi wisata [18]. Hal ini menunjukkan bahwa analisis sentimen wisatawan terhadap produk dan layanan wisata di Pulau Komodo dan Pulau Rinca perlu dilakukan untuk menghasilkan rekomendasi peningkatan kualitas produk dan layanan yang sesuai dengan preferensi wisatawan dari waktu ke waktu.

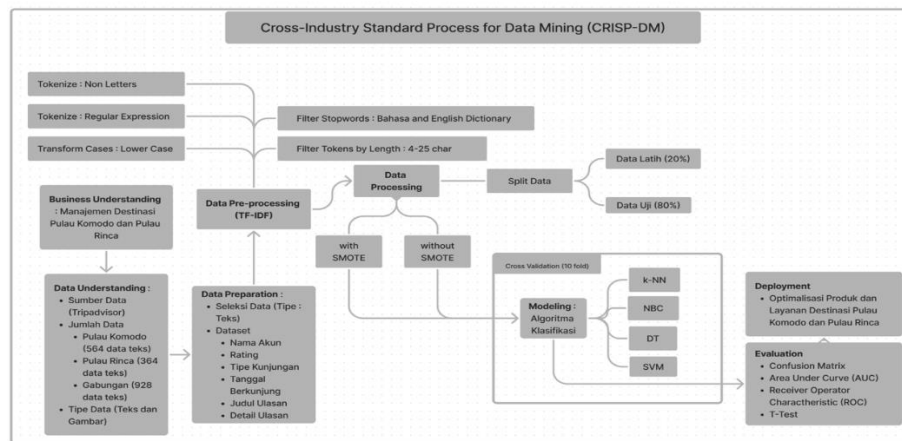
Beberapa penelitian terdahulu menunjukkan bahwa hasil klasifikasi dari algoritma yang digunakan ditentukan oleh data yang telah dikumpulkan. Widiastuti menunjukkan bahwa hasil pengujian algoritma akan menunjukkan akurasi, presisi, recall klasifikasi kelas sentimen dan positif, serta nilai Area Under Curve (AUC) yang berbeda-beda [19]. Disisi lain, Murnawan dan Sinaga menunjukkan bahwa algoritma dengan performa terbaik dapat direkomendasikan sebagai model yang dapat digunakan untuk klasifikasi sentimen wisatawan dalam pemeringkatan destinasi wisata atau produk dan layanan populer [20]. Disisi lain, Wilianto et al. menunjukkan bahwa algoritma yang pernah digunakan dalam kajian analisis sentimen wisatawan ialah Naïve Bayes Classifier (NBC) [21]. Adapun, Noori menunjukkan bahwa algoritma k-NN dan DT juga dapat digunakan dalam klasifikasi sentimen konsumen terhadap produk dan layanan yang ditawarkan, termasuk di sektor pariwisata [22]. Selain itu, Sukiakhi et al. menunjukkan bahwa algoritma SVM juga dapat digunakan dalam klasifikasi data teks untuk

mengidentifikasi dan menganalisis sentimen wisatawan terhadap objek wisata [23]. Hal ini menunjukkan bahwa algoritma k-Nearest Neighbor (k-NN), Naïve Bayes Classifier (NBC), Support Vector Machine (SVM), dan Decision Tree (DT) dapat digunakan dalam mengklasifikasi sentimen wisatawan. Dengan demikian, penelitian ini akan menggunakan dan membandingkan masing-masing algoritma agar dapat diketahui algoritma dengan performa terbaik dalam mengklasifikasi sentimen wisatawan terhadap produk dan layanan wisata di Pulau Komodo dan Pulau Rinca, Manggarai, Nusa Tenggara Timur, Indonesia.

## 2. METODOLOGI PENELITIAN

### 2.1 Cross-Industry Standard Process for Data Mining (CRISP-DM)

Metode yang digunakan dalam analisis sentimen pengunjung Pulau Komodo dan Pulau Rinca di website Tripadvisor ialah Cross Industry Standard Process for Data Mining (CRISP-DM) yang terbagi menjadi enam tahapan sebagai berikut business understanding, data understanding, data preparation, modeling, evaluation dan deployment. Pada tahap business understanding, dilakukan studi literatur tentang sistem manajemen destinasi wisata, manajemen pemasaran destinasi berbasis website, segmentasi dan preferensi wisatawan terhadap produk dan layanan di destinasi wisata; tahap data understanding, dilakukan pengujian sistem validasi data ulasan oleh member website Tripadvisor, screening data ulasan berdasarkan nama akun, rating, tipe kunjungan, tanggal berkunjung, judul ulasan, dan detail ulasan, kemudian melakukan proses scraping data teks menggunakan aplikasi webharvy; tahap data preparation, dilakukan proses seleksi data yang akan dibersihkan (tokenize, transform cases, filter stopwords, filter tokens by length); tahap data processing, dilakukan pembagian data latih (20%) dan data uji (80%) dari 654 data ulasan Pulau Komodo, 364 data ulasan Pulau Rinca, dan 928 total gabungan data ulasan Pulau Komodo dan Pulau Rinca yang telah dihimpun. Selanjutnya, dilakukan pengujian algoritma menggunakan operator SMOTE Upsampling dan tanpa menggunakan SMOTE Upsampling untuk mengidentifikasi perbandingan nilai Area Under Curve (AUC); tahap modeling, dilakukan klasifikasi dataset menggunakan operator cross validation pada algoritma k-Nearest Neighbor (k-NN), Naïve Bayes Classifier (NBC), Decision Tree (DT), Support Vector Machine (SVM); tahap evaluation, dilakukan analisis nilai confusion matrix, f-measure, Area Under Curve (AUC), dan Receiver Operating Characteristic (ROC), and t-Test; tahap deployment, diberikan rekomendasi model dan luaran hasil analisis sentimen untuk optimalisasi produk dan layanan destinasi wisata Pulau Komodo dan Pulau Rinca. Adapun, alur dari keseluruhan tahapan dapat dilihat pada gambar 1 berikut ini.



**Gambar 1.** Tahapan Penelitian sesuai CRISP-DM Framework

Gambar 1 merupakan keseluruhan tahapan penelitian berdasarkan kerangka kerja CRISP-DM dalam menganalisis analisis sentimen pengunjung Pulau Komodo dan Pulau Rinca di website Tripadvisor menggunakan algoritma k-Nearest Neighbor (k-NN), Naïve Bayes Classifier (NBC), Decision Tree (DT), Support Vector Machine (SVM). Masing-masing algoritma memiliki keunggulan dan kelemahan sehingga perlu disesuaikan dengan karakteristik data yang akan dikelola untuk menghasilkan model yang sesuai. Hasil evaluasi algoritma dengan performa terbaik menunjukkan bahwa data tersebut dapat digunakan sebagai dasar untuk menghasilkan rekomendasi bagi pengelola destinasi wisata dalam mengoptimalkan produk dan layanan di Pulau Komodo dan Pulau Rinca, Manggarai, Nusa Tenggara Timur (NTT), Indonesia.

### 2.2 Pembobotan Kata (Term Frequency-Inverse Document Frequency/TF-IDF)

Pembobotan kata merupakan proses pemberian nilai pada setiap kata yang telah melewati tahap Pre-Processing. Penelitian ini menggunakan metode Term Frequency-Inverse Document Frequency (TF-IDF) dengan persamaan (1) dan (2) berikut :

$$IDF(w) = \log\left(\frac{N}{DF(w)}\right) \quad (1)$$

$$W_{ij} = TF_{ij} \times \log\left(\frac{D}{DF_{ij}}\right) \quad (2)$$

Kosasih dan Alberto menunjukkan bahwa proses pembobotan kata dapat dilakukan dengan beberapa tahapan sebagai berikut : pertama, menghitung jumlah Term Frequency (TF) tiap kata, dimana kalimat yang telah dipisah menjadi kata akan diberi nilai dan setiap kata yang muncul akan diberi nilai 1; kedua, menghitung jumlah Document Frequency (DF) tiap kata dengan cara menjumlahkan nilai TF pada tiap kata; ketiga, menghitung jumlah Inverse Document Frequency (IDF) yang ditunjukkan pada persamaan (1); keempat, menghitung bobot (Weight) pada tiap kata yang diperoleh dari hasil perkalian nilai TF dengan IDF sebagaimana persamaan (2) [24]. Selanjutnya, hasil pembobotan kata akan digunakan sebagai input untuk algoritma k-Nearest Neighbor (k-NN), Naïve Bayes Classifier (NBC), Decision Tree (DT), Support Vector Machine (SVM).

### 2.3 Penyeimbang Data (Synthetic Minority Oversampling Technique/SMOTE)

Barro et al. berpendapat bahwa ketidakseimbangan data akan terjadi apabila jumlah objek di suatu kelas data memiliki kuantitas yang lebih tinggi dibandingkan dengan kelas lain, dimana kelas data yang objeknya lebih banyak disebut kelas mayor sedangkan yang lain disebut minor [25]. Disisi lain, Kurniawati menekankan bahwa pengolahan algoritma yang tidak mempertimbangkan ketidakseimbangan data cenderung menitikberatkan kelas mayor dan bukan kelas minor, oleh sebab itu diperlukan teknik SMOTE yang menggunakan metode oversampling untuk memperbanyak pengamatan secara acak dengan menambah jumlah data kelas minor (data buatan) agar setara dengan kelas mayor [26]. Adapun, data buatan atau sintesis tersebut dibuat berdasarkan k-tetangga terdekat (k-Nearest Neighbor). Pembangkit data buatan yang berskala numerik diukur jarak kedekatannya dengan jarak euclidean sedangkan data kategorik berdasarkan kelas minor yang peubahnya berskala kategorik, dilakukan dengan rumus Value Difference Metric (VDM) yaitu :

$$\Delta(x, y) = w_x w_y \sum_{i=1}^N \delta(x_i y_i)^r \quad (3)$$

Persamaan (3) merupakan proses untuk membangkitkan data numerik. Dimana  $\Delta(x, y)$  adalah jarak antara amatan  $x$  dengan  $y$ , sementara  $w_x w_y$  merupakan bobot amatan (dapat diabaikan),  $N$  merupakan banyaknya pebuah penjelas,  $r$  bernilai 1 (jarak manhattan) atau 2 (jarak euclidean), serta  $\delta(x_i y_i)^r$  jarak antar kategori. Adapun, proses pembangkit data buatan (sintesis) untuk data numerik dilakukan dengan menghitung perbedaan antar vektor utama dengan k-tetangga terdekatnya, kalikan perbedaan dengan angka yang diacak diantara 0 dan 1, kemudian tambahkan perbedaan tersebut ke dalam nilai utama pada vektor utama asal sehingga diperoleh vektor utama yang baru. Selanjutnya, pembangkit data kategorik dapat dilakukan melalui persamaan (4) sebagai berikut.

$$\delta(V_1 V_2) = \sum_{i=1}^n \left| \frac{C_{1i}}{C_1} - \frac{C_{2i}}{C_2} \right|^K \quad (4)$$

Dimana,  $\delta(V_1 V_2)$  merupakan jarak antara nilai  $V_1$  dan  $V_2$  sedangkan  $C_{1i}$  merupakan banyaknya  $V_1$  yang termasuk kelas I, dan  $C_{2i}$  merupakan banyaknya  $V_2$  yang termasuk kelas I. Sementara itu,  $i$  merupakan banyaknya kelas,  $C_1$  banyaknya nilai 1 terjadi,  $C_2$  banyaknya nilai 2 terjadi,  $n$  merupakan banyaknya kategori, dan  $k$  merupakan konstansi. Proses pembangkitan data buatan (sintesis) untuk data kategori dilakukan dengan memilih mayoritas antara vektor utama yang dipertimbangkan dengan k-tetangga terdekatnya untuk nilai nominal, jika nilai sama maka akan dipilih secara acak. Selanjutnya, nilai tersebut dijadikan data contoh kelas buatan yang baru.

### 2.4 Algoritma k-Nearest Neighbor

k-Nearest Neighbor (k-NN) merupakan salah satu metode untuk melakukan klasifikasi berdasarkan data pembelajaran dengan tujuan mengklasifikasikan objek tertentu sebagaimana atribut dan sample dengan jarak yang terdekat dengan objek [27]. Algoritma k-NN menggunakan perhitungan jarak euclidean (euclidean distance) dengan rumus sebagai berikut :

$$d = \sqrt{\sum_{i=1}^n (x_{training}^i - x_{testing}^i)^2} \quad (5)$$

Dimana  $d$  merupakan jarak euclidean,  $x_{training}^i$  merupakan data training,  $x_{testing}^i$  merupakan data testing,  $i$  merupakan variabel data, dan  $n$  merupakan banyaknya data. Setelah melakukan perhitungan jarak antara data testing dan data training, maka data akan diurutkan dari yang terkecil ke yang terbesar. Selanjutnya, nilai  $k$  digunakan untuk batas tetangga terdekat, berdasarkan nilai  $k$  maka dapat diperoleh data mayoritas sebagai hasil prediksi.

### 2.5 Algoritma Naïve Bayes Classifier

Naïve Bayes Classifier (NBC) memiliki keunggulan tersendiri dimana data diklasifikasi dengan probabilitas sederhana yang mengaplikasikan teorema bayes dengan asumsi ketidaktergantungan (independen) yang tinggi [28]. Penelitian ini didasarkan pada jumlah dataset yang digunakan untuk metode yang mempunyai performansi

cepat dan akurat dalam pengklasifikasian. Naïve Bayes Classifier hanya membutuhkan data latih (training data) yang relatif kecil untuk menentukan estimasi parameter yang dibutuhkan dalam proses klasifikasi. Pada tahap klasifikasi, nilai kategori ditentukan dari data berdasarkan term yang muncul menggunakan persamaan berikut.

$$P(X_k|Y) = \frac{P(Y|X_k)}{\sum_i P(Y|X_i)} \quad (6)$$

Dimana, keadaan posterior (Probabilitas  $X_k$  di dalam  $Y$ ) dapat dihitung dari keadaan prior (Probabilitas  $Y$  di dalam  $X_k$ ) dibagi dengan jumlah dari semua probabilitas  $Y$  di dalam semua  $X_i$ . Dalam konteks penelitian ini, data teks yang diperoleh dari website Tripadvisor diklasifikasi menggunakan persamaan berikut

$$P(v_1|C=c) = \frac{\text{CountTerms}(v_1, \text{docsv}(c))}{\text{AllTerms}(\text{docs}(c))} \quad (7)$$

Dimana  $v_1$  merupakan salah satu suku kata yang muncul dalam ulasan pengguna website Tripadvisor terhadap kualitas produk dan layanan di Destinasi Wisata Candi Borobudur. Sedangkan,  $\text{CountTerms}(v_1, \text{docsv}(c))$  merujuk pada jumlah kemunculan suatu kata berlabel  $c$  (“positif” atau “negatif”). Adapun,  $\text{AllTerms}(\text{docs}(c))$  merujuk pada jumlah semua kata berlabel  $c$  yang ada pada dataset. Untuk menghindari adanya nilai nol pada probabilitas maka diimplementasikan laplace smoothing, untuk mengurangi probabilitas dari hasil yang terobservasi, dan jua meningkatkan probabilitas hasil yang belum terobservasi. Dengan demikian, persamaan yang digunakan ialah sebagai berikut :

$$P(v_1|C=c) = \frac{\text{CountTerms}(v_1, \text{docsv}(c))+1}{\text{AllTerms}(\text{docs}(c))+|V|} \quad (8)$$

Dimana  $|V|$  merujuk pada jumlah semua kata dalam data ulasan yang ada di dataset. Dengan demikian, proses klasifikasi data ulasan akan menunjukkan kata dengan nilai tertinggi sebagai representasi perhatian pengulas terhadap produk dan jasa layanan pariwisata di Destinasi Candi Borobudur melalui website Tripadvisor.

## 2.6 Algoritma Decision Tree

Decision Tree (DT) merupakan salah satu metode klasifikasi populer yang sering digunakan secara praktis, model prediksi yang menggunakan struktur pohon untuk mencari dan membuat keputusan, serta memecahkan masalah dengan mempertimbangkan berbagai faktor di dalam lingkup masalah tersebut [29]. Decision Tree memiliki beberapa algoritma salah satunya Iterative Dychotomizer version (ID3), yaitu model klasifikasi yang berupa pohon keputusan secara top-down dengan cara kerja mengevaluasi semua atribut menggunakan suatu ukuran statistik berupa information gain untuk mengukur efektifitas suatu atribut dalam mengklasifikasi sample data. Dalam algoritma ini, dibutuhkan nilai entropy dan gain, dimana entropy merupakan parameter untuk mengukur jumlah keberagaman atau keberadaan dalam sebuah himpunan data, sedangkan gain merupakan perolehan informasi sebagai ukuran efektifitas suatu atribut. Berikut adalah persamaan untuk mendapatkan nilai entropy dan gain.

$$\text{Entropy}(S) = \sum_{i=1}^n -p_i \times \log_2 p_i \quad (9)$$

$$\text{Gain}(S, A) = S - \sum_{i=1}^n \frac{|S_i|}{|S|} \times S_i \quad (10)$$

Dimana  $S$  merupakan nilai Entropy,  $p_i$  jumlah yang memiliki nilai positif atau negatif pada kumpulan data untuk sifat tertentu. Disisi lain,  $\text{Gain}(S, A)$  adalah hasil informasi yang berasal dari luaran data yang dikelompokkan sesuai dengan atribut  $A$ . Selanjutnya,  $S_i$  adalah subset dari nilai entropy yang mempunyai nilai  $i$ . Adapun,  $S$  adalah subset dari nilai Entropy. Puspita & Widodo berpendapat bahwa algoritma Decision Tree memiliki keunggulan dan kekurangan, dimana keunggulannya ialah konsep yang jelas dan mudah dipahami serta diimplementasikan menggunakan algoritma rekursif, sedangkan kekurangannya ialah tidak dapat diaplikasikan pada himpunan data yang sangat besar dan mudah mengalami overfit karena proses pelatihan greedy [30]. Dengan demikian perlu dilakukan pengujian algoritma SVM untuk membandingkan performa algoritma yang lebih baik.

## 2.7 Algoritma Support Vector Machine

Support Vector Machine (SVM) merupakan salah satu algoritma yang dapat digunakan untuk klasifikasi data menggunakan hyperplane [31]. Karim menjelaskan bahwa konsep SVM menitikberatkan pada risk minimization, yaitu estimasi fungsi dengan cara meminimalisir batas dari generalization error, sehingga SVM mampu mengatasi overfitting [32]. Adapun, fungsi regresi dari metode SVM adalah sebagai berikut.

$$f(x) = w^T \varphi(x) + b \quad (11)$$

Dimana  $w$  merupakan vector pembobot,  $\varphi(x)$  merupakan sebuah fungsi yang memetakan  $x$  ke dalam suatu dimensi, dan  $b$  merupakan faktor bias. Selanjutnya, Nida et al., menunjukkan bahwa SVM memiliki kelebihan dalam generalisasi data yang tinggi serta mampu menghasilkan model klasifikasi yang baik meskipun dilatih dengan data yang relatif sedikit. Meskipun demikian, sangat sulit diaplikasikan untuk himpunan data dengan sampel dan dimensi yang besar [33]. Hal ini menunjukkan bahwa SVM mampu menghasilkan performa yang baik

meskipun dengan jumlah data yang relatif sedikit. Dengan demikian, perlu dilakukan pengujian algoritma SVM dengan dataset ulasan pengunjung Pulau Komodo dan Pulau Rinca.

### 2.8 Evaluasi Klasifikasi (Confusion Matrix)

Kasanah et al. menunjukkan bahwa Evaluasi klasifikasi didasarkan pada pengujian pada objek yang benar dan objek yang salah. Validasi digunakan untuk menentukan jenis model yang terbaik melalui confusion matrix sebagai informasi mengenai hasil klasifikasi actual yang dapat diprediksi oleh suatu sistem melalui nilai akurasi, presisi, dan recall, melalui persamaan berikut.

$$\text{Akurasi} = \frac{TP+TN}{TP+FP+TN+FN} \tag{12}$$

$$\text{Presisi/Specificity} = \frac{TP}{TP+FP} \tag{13}$$

$$\text{Recall/Sensitivity} = \frac{TP}{TP+FN} \tag{14}$$

$$f - \text{measure} = \frac{2 \times (\text{Presisi} \times \text{recall})}{\text{presisi} + \text{recall}} \tag{15}$$

Hairani et al. memberikan penjelasan tentang confusion matrix dari akurasi, presisi, recall dan f-measure : akurasi adalah ketepatan sistem dalam melakukan proses klasifikasi dengan benar; presisi atau sensitivity adalah rasio jumlah dokumen yang relevan dengan total jumlah dokumen yang ditemukan pada sistem klasifikasi; recall atau specificity adalah rasio jumlah dokumen yang ditemukan kembali oleh sistem klasifikasi dengan total jumlah dokumen yang relevan; f-measure adalah metrik evaluasi yang populer untuk menangani masalah imbalance class dengan mengombinasikan recall/sensivitas dan presisi sehingga menghasilkan metrik yang efektif untuk mencari Kembali informasi dalam himpunan yang tidak seimbang [34]. Dengan demikian, evaluasi performa algoritma terbaik dapat direkomendasikan sebagai model yang relevan dengan dataset untuk memperoleh luaran analisis sentimen yang tergolong fit atau sesuai.

## 3. HASIL DAN PEMBAHASAN

Pembahasan penelitian ini terbagi menjadi dua bagian yakni : pertama, analisis sentimen Pulau Komodo dan Pulau Rinca; evaluasi performa algoritma dalam klasifikasi data sentimen. Secara spesifik, pembahasan dalam penelitian ini menekankan pada evaluasi performa algoritma k-NN, NBC, SVM, dan DT berdasarkan nilai akurasi, presisi dan recall ketika mengklasifikasi data sentimen wisatawan untuk menghasilkan rekomendasi pengembangan sistem informasi yang meningkatkan kualitas produk dan layanan di destinasi wisata Pulau Komodo dan Pulau Rinca, Manggarai, Nusa Tenggara Timur, Indonesia.

### 3.1 Analisis Sentimen Pulau Komodo dan Pulau Rinca

Pulau Komodo dan Pulau Rinca merupakan bagian dari destinasi wisata yang menarik wisatawan mancanegara. Ziku berpendapat bahwa pengembangan pariwisata Pulau Komodo dan Pulau Rinca didukung oleh masyarakat lokal yang terejawantahkan dalam partisipasi di kegiatan konservasi dan bisnis pariwisata [35]. Disisi lain, Idris et al. menunjukkan bahwa pengembangan pariwisata di Pulau Komodo memiliki manfaat yang positif pada Pendapatan Asli Daerah (PAD), hal tersebut juga dapat dilihat dari pendapatan masyarakat, terpenuhinya kebutuhan keluarga, serta kebutuhan pendidikan dan fasilitas kesehatan yang layak [36]. Bahkan, Widarto dan Mubasyira berpendapat bahwa wacana penutupan sementara destinasi wisata Pulau Komodo menimbulkan dilema bagi masyarakat lokal, karena sebagian masyarakat lokal memiliki sumber matapencaharian dari aktivitas wisata. Disisi lain, wisatawan mancanegara yang mengunjungi Pulau Komodo memiliki sentimen tersendiri terkait dengan produk dan layanan dalam bentuk atraksi, jasa akomodasi dan aksesibilitas bagi wisatawan mancanegara maupun wisatawan domestik untuk mengunjungi Pulau Komodo dan Pulau Rinca. Hal ini menunjukkan bahwa analisis sentimen wisatawan yang berkunjung ke Pulau Komodo maupun Pulau Rinca perlu dilakukan untuk menghasilkan sejumlah rekomendasi bagi pengelola dalam mengoptimalkan produk dan layanan wisata. Mempertimbangkan hal tersebut maka, penelitian ini mengumpulkan data teks dari hasil ulasan wisatawan Pulau Komodo dan Pulau Rinca di halaman website Tripadvisor, sebagaimana Tabel 1 berikut ini.

**Tabel 1.** Jumlah Data Teks Ulasan Wisatawan Pulau Rinca dan Pulau Komodo

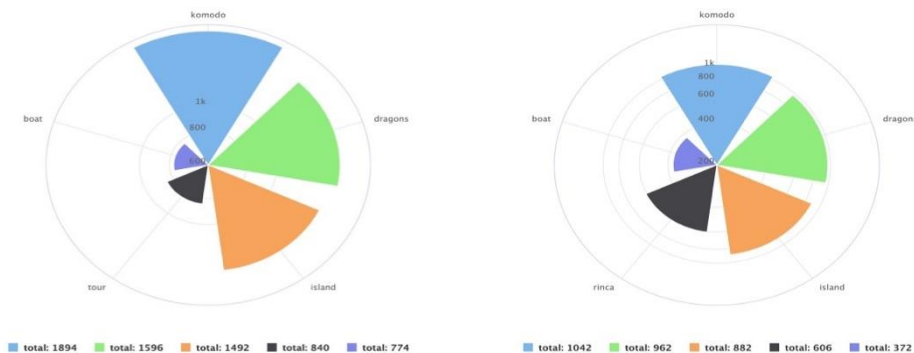
Dataset	Jumlah Data (Sebelum Filter)	Jumlah Data (Setelah Filter)	Link Tripadvisor
Pulau Komodo	943	564	<a href="https://www.tripadvisor.com/Attraction_Review-g325592-d3247649-Reviews-Komodo_Island-Komodo_National_Park_Komodo_East_Nusa_Tenggara.html">https://www.tripadvisor.com/Attraction_Review-g325592-d3247649-Reviews-Komodo_Island-Komodo_National_Park_Komodo_East_Nusa_Tenggara.html</a>
Pulau Rinca	719	364	<a href="https://www.tripadvisor.com/Attraction_Review-g1820452-d3253186-Reviews-Rinca_Island-Rinca_Komodo_National_Park_Komodo_East_Nusa_Tenggara.html">https://www.tripadvisor.com/Attraction_Review-g1820452-d3253186-Reviews-Rinca_Island-Rinca_Komodo_National_Park_Komodo_East_Nusa_Tenggara.html</a>

Tabel 1 merupakan hasil pengumpulan data teks dari kolom ulasan Website Tripadvisor untuk destinasi wisata Pulau Komodo dan Pulau Rinca. Pengumpulan data ulasan menggunakan aplikasi webharvy untuk berdasarkan konfigurasi nama akun, rating, tipe kunjungan, tanggal berkunjung, judul ulasan, dan detail ulasan. Berdasarkan hasil pengumpulan data, diperoleh 564 data dari 943 data ulasan Pulau Komodo dan 364 data dari 719 data ulasan dengan filter data ulasan dalam bahasa Inggris. Selanjutnya Data teks yang terkumpul diseleksi dan dibersihkan menggunakan aplikasi Rapidminer melalui operator process documents from data yang didalamnya terdapat operator untuk membersihkan dan menyiapkan data teks seperti tokenize, transform cases, filter tokens by length, dan filter stopwords. berdasarkan konfigurasi sebagai berikut : pertama, tokenize 1 (non-letters), tokenize 2 (regular expression [-!\"#\$%&'()\*+,-./:;<=>?@\[\]\\_`{|}~]); kedua, transform cases (lower case); ketiga, filter tokens by length (min char =4 & max chars=25); keempat, filter stopwords (dictionary English). Selain itu, terdapat operator remove duplicate, untuk memastikan tidak ada data ulasan yang duplikat, kemudian digunakan operator sentiment extract untuk pembobotan kata sehingga mendapatkan string score sebagaimana tabel berikut ini.

**Tabel 2.** Proses Ekstrak Sentimen dari Ulasan Wisatawan

Review					
Visiting this island was the absolute <b>worst</b> tourist experience I've ever had. The island is filthy, the smells are revolting <b>like</b> raw sewage. The dragon we saw was either drugged or <b>dead</b> . We were told by our guide it would be a 40 minute trek to see dragons . After seeing one <b>pathetic</b> animal who was too wretched to even photograph our guide announced that we were very <b>lucky</b> and headed off to rip off his next <b>victims</b> . Indonesia's minister for tourist should be <b>ashamed</b> . Don't go					
String Scoring					
worst (-0.79)	like (0.38)	dead (-0.85)	pathetic (-0.69)	lucky (0.46)	victims (-0.33) ashamed (-0.54)
Sentiment	Score	Negativity	Positivity	Uncovered Tokens	Tokens
Negatif	-2,4	3,2	0,8	89	96
Review					
As the part of New7Wonders of Nature, Komodo Island is very <b>beautiful</b> . When you met the dragon, they look so tempting. It is <b>good</b> to start your day by tracking the forest of Komodo National Park and end up by saw the Komodo <b>relaxing</b> their morning time at the beach. I <b>guarantee</b> you will be very <b>excited</b> when you see this exotic dragon. Meanwhile, you can also <b>enjoy</b> the <b>clear</b> sea water when you are sailing from Labuan Bajo to the Komodo Island.					
String Scoring					
beautiful (0.74)	good (0.49)	relaxing (0.56)	guarantee (0.26)	excited (0.36)	enjoy (0.56) clear (0.41)
Sentiment	Score	Negativity	Positivity	Uncovered Tokens	Tokens
Positif	3,4	0	3,4	83	90

Tabel 2 merupakan hasil ekstrak sentimen dari ulasan wisatawan dimana operator extract sentiment di aplikasi Rapidminer akan mengidentifikasi dan melakukan seleksi kata, memberikan bobot, kemudian mengakumulasi bobot tersebut sehingga dapat diklasifikasikan berdasarkan sentimen positif dan sentimen negatif. Salah satu kekurangan dari operator extract sentiment ialah pembobotan (string scoring) hanya diberikan berdasarkan diksi atau kosa kata yang mengandung makna positif atau negatif. Sementara itu, analisis sentimen merupakan proses identifikasi dan klasifikasi berdasarkan makna teks secara keseluruhan, bukan parsial. Hal ini menjadi tantangan bagi pengembangan sistem aplikasi dalam konteks data mining agar dapat mempertimbangkan makna keseluruhan dari kalimat yang diulas oleh wisatawan tentang destinasi wisata Pulau Komodo dan Pulau Rinca, entah itu ekspresi kekaguman, sarkasme, sindiran, protes, komplain dari wisatawan. Meskipun demikian, hasil TF-IDF dan proses pembobotan kata menggunakan operator wordlist dan sortir data dapat divisualisasikan dalam bentuk polar chart 5 kata teratas yang paling sering muncul dalam ulasan wisatawan di Pulau Komodo dan Pulau Rinca sebagai berikut.



**Gambar 2** Polar Chart Top 5 Words dari Data Ulasan Wisatawan

Gambar 2 merupakan visualisasi data ulasan dari pengunjung destinasi wisata Pulau Komodo dan Pulau Rinca dalam bentuk polar chart. Berdasarkan hasil klasifikasi data menggunakan metode pembobotan kata (TF-IDF) melalui operator ekstrak sentimen dapat diketahui lima kata teratas yang paling sering muncul dalam ulasan wisatawan terhadap Pulau Komodo ialah sebagai berikut : komodo (1894), dragons (1596), island (1492), tour (840), boat (774). Sementara itu, lima kata teratas yang paling sering muncul dalam ulasan wisatawan terhadap Pulau Rinca ialah sebagai berikut : komodo (1042), dragons (962), island (882), rinca (606), boat (372). Berdasarkan hasil perankingan kata, dapat diinterpretasikan bahwa kesan pengunjung Pulau Komodo dan Pulau Rinca memberikan penekanan pada atraksi (Pulau Komodo, Pulau Rinca, dan Komodo), akomodasi dan aksesibilitas yaitu penginapan dan jasa transportasi laut (tour dan boat). Apabila dihubungkan dengan kondisi eksisting bisnis pariwisata, kondisi eksisting pariwisata Pulau Komodo dan Pulau Rinca dapat dianalisis berdasarkan elemen 3-A (Atraksi, Akomodasi dan Aksesibilitas) untuk merekomendasikan pengembangan sistem informasi yang dapat mengoptimalkan produk dan layanan pariwisata Pulau Komodo dan Pulau Rinca, sebagaimana tabel berikut.

**Tabel 3.** Analisis Kondisi Eksisting dan Rekomendasi Pengembangan Sistem

3-A & Top 5 Words	Analisis Kondisi Eksisting 3-A di Pulau Komodo dan Pulau Rinca	Rekomendasi Pengembangan Sistem untuk Optimalisasi Produk dan Layanan Pariwisata Pulau Komodo dan Pulau Rinca.
Atraksi (Rinca, Komodo, Dragons, Island )	<p>Kondisi eksisting destinasi wisata Pulau Komodo dan Pulau Rinca dikemas dalam satu paket perjalanan wisata oleh pelaku usaha perjalanan wisata.</p> <p>Komodo menjadi daya tarik utama wisatawan mancanegara, sehingga perlu dipertimbangkan keberlanjutan ekosistem</p> <p>Kolaborasi berbagai pemangku kepentingan dalam pengembangan destinasi wisata lokal</p>	<p>Mengembangkan sistem informasi manajemen destinasi untuk pemetaan dan pengendalian/monitoring ekosistem di Pulau Komodo dan Pulau Rinca.</p> <p>Mendokumentasikan data kunjungan wisatawan mancanegara untuk pengendalian daya dukung destinasi serta dasar perumusan kebijakan yang tepat guna.</p>
Akomodasi (Boat dan Tour)	<p>Fasilitas dalam moda transportasi laut telah dimodifikasi sehingga memiliki fasilitas menginap sebagai jasa akomodasi.</p> <p>Wisatawan mancanegara menggunakan jasa akomodasi dan transportasi dalam satu paket perjalanan wisata.</p>	<p>Mengembangkan sistem informasi terintegrasi berbasis website untuk pemasaran produk dan layanan wisata berdasarkan klasifikasi sarana dan prasarana (standar-premium), sehingga mendistribusikan konsumen secara kompetitif untuk keberlanjutan ekonomi-pariwisata di Pulau Komodo, Pulau Rinca, dan destinasi wisata sekitar.</p>
Aksesibilitas (Boat)	<p>Persaingan bisnis perjalanan wisata berdasarkan klasifikasi produk dan layanan (standar-premium) moda transportasi dan akomodasi.</p>	

Tabel 3 merupakan hasil interpretasi lima kata teratas dari hasil TF-IDF dan ekstrak sentimen yang dianalisis berdasarkan kondisi eksisting Atraksi, Akomodasi dan Aksesibilitas (3-A) di Pulau Komodo dan Pulau Rinca untuk merekomendasikan sistem informasi yang relevan dan kontekstual, dalam upaya optimalisasi produk dan layanan pariwisata lokal. Hal ini menunjukkan bahwa pendekatan CRISP-DM mampu menghubungkan proses data mining dengan kajian kepariwisataan untuk menghasilkan sejumlah rekomendasi pengembangan produk dan layanan di sektor pariwisata. Prakoso et al. menunjukkan bahwa pengusaha jasa perjalanan wisata perlu mengembangkan sistem informasi reservasi paket wisata berlayar menuju Pulau Komodo, agar memberikan kemudahan informasi bagi wisatawan mancanegara, sehingga menarik minat dan meningkatkan kunjungan wisatawan [37]. Selain itu, Chasanah et al. menunjukkan bahwa pengembangan sistem informasi dan produk pariwisata dapat meningkatkan citra destinasi sehingga menarik kunjungan secara berkala [38]. Hal ini menunjukkan bahwa pariwisata di Pulau Komodo dan Pulau Rinca dapat berkembang pesat apabila memiliki citra yang baik. Dengan demikian, rekomendasi dari hasil penelitian dapat ditindaklanjuti untuk mengoptimalkan produk dan layanan pariwisata Pulau Komodo dan Pulau Rinca, Manggarai, Nusa Tenggara Timur, Indonesia.

**3.2 Evaluasi Performa Algoritma dalam Klasifikasi Data Sentimen**

Berdasarkan hasil klasifikasi 564 data ulasan wisatawan Pulau Komodo dan 364 data ulasan wisatawan Pulau Rinca menggunakan algoritma k-NN, NBC, SVM, dan DT berbasis CRISP-DM, dapat diketahui bahwa hasil pengolahan data ulasan wisatawan terhadap produk dan layanan Pulau Komodo, model pengolahan data dengan performa algoritma terbaik ialah Support Vector Machine (SVM) dengan nilai akurasi 99,69%, presisi 100%, recall 99,39%, f-measure 99,69%, Area Under Curve (AUC) 100% dan t-Test 0,958. Sementara, hasil pengolahan data ulasan wisatawan terhadap produk dan layanan Pulau Rinca, model pengolahan data dengan performa algoritma terbaik ialah Support Vector Machine (SVM) dengan nilai akurasi 100%, presisi 100%, recall 100%, f-measure 100%, Area Under Curve (AUC) 100% dan t-Test 0,964. Hal tersebut dapat dilihat pada tabel 4 dan tabel 5 berikut.

**Tabel 4.** Hasil Klasifikasi Algoritma untuk Data Ulasan Pulau Komodo

Pulau Komodo											
Algoritma	TP	F N	FP	T N	Accuracy	Precision	Recall	F-Measure	AUC	t-Test	Keterangan
K-NN	81	81	0	3	50.18%	100%	0.37%	0.73%	86.6%	0.922	with SMOTE
	8	5		0.016							
K-NN	16	4	66	81	92.22%	92.51%	99.51%	95.88%	88.3%	0.922	without SMOTE
				4						0.017	
NBC	81	12	0	80	99.26%	100%	98.53%	99.25%	54.9%	0.956	with SMOTE
	8			6						0.019	
NBC	53	11	29	80	95.56%	96.57%	98.65%	97.59%	51.6%	0.956	without SMOTE
				7						0.020	
DT	68	10	12	80	91.50%	86332%	98.78%	92.10%	91.6%	0.933	with SMOTE
	9		9	8						0.019	
DT	29	7	53	81	93.33%	93.89%	99.14%	96.44%	67.2%	0.933	without SMOTE
				1						0.020	
SVM	81	5	0	81	99.69%	100%	99.39%	99.69%	100%	0.958	with SMOTE
	8			3						0.015	
SVM	44	0	38	81	95.78%	95.58%	100%	97.73%	97.3%	0.958	without SMOTE
				8						0.016	

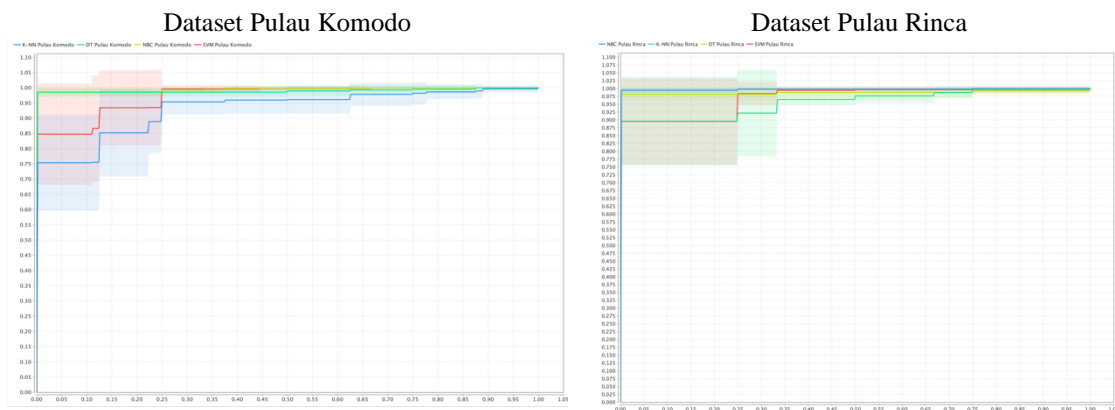
Tabel 4 merupakan hasil klasifikasi data ulasan wisatwan terhadap produk dan layanan Pulau Komodo menggunakan operator SMOTE dan tanpa menggunakan operator SMOTE pada algoritma k-NN, NBC, DT dan SVM. Berdasarkan hasil klasifikasi 564 data ulasan wisatawan Pulau Komodo dapat diketahui bahwa algoritma Support Vector Machine (SVM) dengan operator SMOTE menghasilkan nilai akurasi 99,69%, presisi 100%, recall 99,39%, f-measure 99,69%, Area Under Curve (AUC) 100% dan t-Test 0,958. Hal ini menunjukkan bahwa algoritma SVM memiliki performa yang lebih baik dibandingkan dengan algoritma lainnya, baik k-NN, NBC, dan DT. Adapun, hasil dari klasifikasi algoritma untuk data ulasan Pulau Rinca dapat dilihat pada tabel 5 berikut.

**Tabel 5.** Hasil Klasifikasi Algoritma untuk Data Ulasan Pulau Rinca

Pulau Rinca											
Algoritma	TP	F N	F P	T N	Accuracy	Precision	Recall	F-Measure	AUC	t-Test	Keterangan
K-NN	54	54	0	0	50.31%	unknow n	unknow n	unknow n	85.4%	0.943	with SMOTE
	4	4		0.014							
K-NN	4	0	34	54	94.16%	94.13%	100%	96.97%	91.1%	0.943	without SMOTE
				4						0.015	

Pulau Rinca											
Algoritma	TP	F N	F P	T N	Accurac y	Precisio n	Recall	F- Measu re	AUC	t-Test	Keterang an
NBC	54 4	4	0	54 0	99.63%	100%	99.27%	99.63%	50%	0.969 +/- 0.024	with SMOTE
NBC	25	5	13	53 9	96.91%	97.67%	99.08%	98.36%	50%	0.969 +/- 0.025	without SMOTE
DT	52 5	18	19	52 6	96.60%	96.60%	96.68%	96.59%	95.8 %	0.933 +/- 0.022	with SMOTE
DT	8	9	30	53 5	93.30%	94.72%	98.34%	96.48%	60.4 %	0.933 +/- 0.023	without SMOTE
SVM	54 4	0	0	54 4	100%	100%	100%	100%	100 %	0.964 +/- 0.015	with SMOTE
SVM	17	0	21	54 4	96.39%	96.31%	100%	98.11%	93.8 %	0.964 +/- 0.016	without SMOTE

Tabel 5 merupakan hasil klasifikasi data ulasan wisatawan terhadap produk dan layanan Pulau Rinca menggunakan operator SMOTE dan tanpa menggunakan operator SMOTE pada algoritma k-NN, NBC, DT dan SVM. Berdasarkan hasil klasifikasi 364 data ulasan wisatawan Pulau Rinca dapat diketahui bahwa algoritma Support Vector Machine (SVM) menunjukkan performa yang baik dengan nilai akurasi 100%, presisi 100%, recall 100%, f-measure 100%, Area Under Curve (AUC) 100% dan t-Test 0,964. Apabila ditinjau berdasarkan nilai Receiving Operating Characteristic (ROC) dapat diketahui bahwa selain SVM, algoritma NBC dan DT juga memiliki potensi kinerja yang baik, sebagaimana gambar 3 berikut.



**Gambar 3.** ROC Algoritma k-NN, NBC, DT, dan SVM untuk Dataset Pulau Komodo dan Pulau Rinca

Gambar 3 merupakan kurva Receiving Operating Characteristic (ROC) yang menunjukkan bahwa algoritma NBC dan DT juga memiliki potensi kinerja yang baik dalam klasifikasi data ulasan wisatawan terhadap produk dan layanan wisata di Pulau Komodo dan Pulau Rinca. Dengan demikian, penelitian ini memberikan rekomendasi algoritma SVM sebagai model yang relevan digunakan dalam mengklasifikasi sentimen wisatawan untuk meninjau kondisi eksisting pariwisata di Pulau Komodo dan Pulau Rinca, sehingga dapat membuat perencanaan strategis untuk pengembangan pariwisata yang berkelanjutan.

#### 4. KESIMPULAN

Hasil penelitian ini menekankan pada sejumlah rekomendasi optimalisasi produk dan layanan pariwisata di Pulau Komodo dan Pulau Rinca sebagai berikut : pertama, mengembangkan sistem informasi manajemen destinasi untuk pemetaan dan pengendalian/monitoring ekosistem di Pulau Komodo dan Pulau Rinca; kedua, mendokumentasikan data kunjungan wisatawan mancanegara untuk pengendalian daya dukung destinasi serta dasar perumusan kebijakan yang tepat guna; ketiga, Mengembangkan sistem informasi terintegrasi berbasis website untuk pemasaran produk dan layanan wisata berdasarkan klasifikasi sarana dan prasarana (standar-premium), sehingga mendistribusikan konsumen secara kompetitif untuk keberlanjutan ekonomi-pariwisata di Pulau Komodo, Pulau



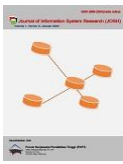
Rinca, dan destinasi wisata sekitar. Selain itu, hasil klasifikasi data menggunakan metode pembobotan kata atau Term Frequency Inverse Document Frequency (TF-IDF) melalui operator ekstrak sentimen menunjukkan bahwa lima kata teratas yang paling sering muncul dalam ulasan wisatawan terhadap Pulau Komodo ialah sebagai berikut : komodo (1894), dragons (1596), island (1492), tour (840), boat (774). Sementara itu, lima kata teratas yang paling sering muncul dalam ulasan wisatawan terhadap Pulau Rinca ialah sebagai berikut : komodo (1042), dragons (962), island (882), rinca (606), boat (372). Adapun, hasil klasifikasi 564 data ulasan wisatawan Pulau Komodo dan 364 data ulasan wisatawan Pulau Rinca menggunakan algoritma k-NN, NBC, SVM, dan DT berbasis CRISP-DM, menunjukkan bahwa Support Vector Machine (SVM) merupakan algoritma dengan performa terbaik dimana nilai akurasi 99,69%, presisi 100%, recall 99,39%, f-measure 99,69%, Area Under Curve (AUC) 100% dan t-Test 0,958. Disisi lain, hasil pengolahan data ulasan wisatawan terhadap produk dan layanan Pulau Rinca, menunjukkan bahwa algoritma Support Vector Machine (SVM) memiliki performa terbaik dengan nilai akurasi 100%, presisi 100%, recall 100%, f-measure 100%, Area Under Curve (AUC) 100% dan t-Test 0,964. Setelah membandingkan performa SVM sebelum dan setelah menggunakan Synthetic Minority Oversampling Technique (SMOTE), dapat diketahui bahwa peforma algoritma menjadi lebih optimal apabila menggunakan operator SMOTE. Dengan demikian, SVM merupakan algoritma yang relevan digunakan sebagai model analisis sentimen wisatawan di Pulau Komodo dan Pulau Rinca berdasarkan CRISP-DM.

## UCAPAN TERIMAKASIH

Terima kasih disampaikan kepada seluruh lembaga yang turut berkontribusi memberikan dukungan hingga penelitian ini dapat selesai. Terima kasih kepada Lembaga Penelitian dan Pengabdian Masyarakat (LPPM), Fakultas Ilmu Administrasi Bisnis dan Ilmu Komunikasi (FIABIKOM), Program Studi Pariwisata, Fakultas Teknik (FT), Program Studi Sistem Informasi, Universitas Katolik Indonesia Atma Jaya.

## REFERENCES

- [1] Anju, Mintasrihardi, and Rohana, "Analysis Of Tourism Business Development In Manggarai Barat Nusa Tenggara Timur ( NTT ) With The Development Of Latest Places," *J. Appl. Bus. Bank.*, vol. 3, no. 2, pp. 84–97, 2021, doi: <https://doi.org/10.31764/jabb.v3i2.8212>.
- [2] E. T. Y. Dima and A. Ketmoen, "Analisis Jumlah Kunjungan Pada Objek Wisata Taman Nasional Komodo Bagi Tingkat Perekonomian di Kabupaten Manggarai Barat," *EkoPem J. Ekon. Pembang.*, vol. 4, no. 4, pp. 120–132, 2022, doi: <https://doi.org/10.32938/jep.v4i4.3306>.
- [3] D. Tulipa, V. Rahmawati, D. Rachmawati, and I. Srianta, "Pembentukan Komunitas Usaha Pengolahan Hasil Laut Dan Hasil Bumi Masyarakat Manggarai Barat-Nusa Tenggara Timur," *J. LeECOM (Leverage, Engag. Empower. Community)*, vol. 4, no. 1, pp. 27–38, 2022, doi: [10.37715/leecom.v4i1.2949](https://doi.org/10.37715/leecom.v4i1.2949).
- [4] A. A. Arifiyanti, M. F. Pandji, and B. Utomo, "Analisis Sentimen Ulasan Pengunjung Objek Wisata Gunung Bromo pada Situs Tripadvisor," *Explor. J. Sist. Inf. dan Telemat.*, vol. 13, no. 1, p. 32, 2022, doi: [10.36448/jsit.v13i1.2539](https://doi.org/10.36448/jsit.v13i1.2539).
- [5] D. I. Afidah, Dairoh, S. F. Handayani, R. W. Pratiwi, and S. I. Sari, "Sentimen Ulasan Destinasi Wisata Pulau Bali Menggunakan Bidirectional Long Short Term Memory," *MATRIK J. Manajemen, Tek. Inform. dan Rekayasa Komput.*, vol. 21, no. 3, pp. 607–618, 2022, doi: [10.30812/matrik.v21i3.1402](https://doi.org/10.30812/matrik.v21i3.1402).
- [6] M. Rohini and M. S. Eastaff, "Data Mining Knowledge Discovery and Its Application," *Int. J. Comput. Sci. Eng. Technol.*, vol. 7, no. 11, pp. 10–12, 2021, [Online]. Available: [https://www.researchgate.net/publication/267338458\\_Secure\\_Messaging\\_System\\_using\\_ZKP](https://www.researchgate.net/publication/267338458_Secure_Messaging_System_using_ZKP)
- [7] F. A. Rahman, M. I. Desa, A. Wibowo, and N. A. Haris, "Knowledge discovery database (KDD)-data mining application in transportation," in *International Conference on Electrical Engineering, Computer Science and Informatics (EECSI)*, 2014, vol. 1, pp. 116–119. doi: [10.11591/eecsi.1.357](https://doi.org/10.11591/eecsi.1.357).
- [8] N. L. W. S. R. Ginantra, C. P. Yanti, G. D. Prasetya, I. B. G. Sarasvandana, and I. K. A. G. Wiguna, "Analisis Sentimen Ulasan Villa di Ubud Menggunakan Metode Naive Bayes, Decision Tree, dan k-NN," *Janapati*, vol. 11, no. 3, pp. 205–216, 2022.
- [9] Y. M. W. Wahyu, A. R. Berto, and E. Murwani, "Analisis Sentimen Jaringan Pesan Kolom Komentar Video Wonderful Indonesia 2022 Jagad Jawi yang Dipengaruhi Budaya," *Avant Garde J. Ilmu Komun.*, vol. 10, no. 2, pp. 201–216, 2022.
- [10] D. Arsa, I. Weni, and A. Fahreza, "Analisis Sentimen Terhadap Pariwisata di Masa Covid-19 Menggunakan Naive Bayes," *J. Telemat.*, vol. 17, no. 1, pp. 49–54, 2022.
- [11] Y. A. Singgalen, "Sentiment Analysis on Customer Perception towards Products and Services of Restaurant in Labuan Bajo," *J. Inf. Syst. Informatics*, vol. 4, no. 3, pp. 511–523, 2022, doi: [10.51519/journalisi.v4i3.276](https://doi.org/10.51519/journalisi.v4i3.276).
- [12] H. J. Christanto and Y. A. Singgalen, "Sentiment Analysis on Customer Perception towards Products and Services of Restaurant in Labuan Bajo," *J. Inf. Syst. Informatics*, vol. 4, no. 3, pp. 511–523, 2022, doi: [10.51519/journalisi.v4i3.276](https://doi.org/10.51519/journalisi.v4i3.276).
- [13] Y. A. Singgalen, "Analisis Sentimen Wisatawan Melalui Data Ulasan Candi Borobudur di Tripadvisor Menggunakan Algoritma Naive Bayes Classifier," *Build. Informatics, Technol. Sci.*, vol. 4, no. 3, pp. 1343–1352, 2022, doi: [10.47065/bits.v4i3.2486](https://doi.org/10.47065/bits.v4i3.2486).
- [14] Y. A. Singgalen, "Analisis Performa Algoritma NBC , DT , SVM dalam Klasifikasi Data Ulasan Pengunjung Candi Borobudur Berbasis CRISP-DM," *Build. Informatics, Technol. Sci.*, vol. 4, no. 3, pp. 1634–1646, 2022, doi: [10.47065/bits.v4i3.2766](https://doi.org/10.47065/bits.v4i3.2766).
- [15] Y. A. Singgalen, "Pemilihan Metode dan Algoritma dalam Analisis Sentimen di Media Sosial : Systematic Literature Review," *J. Inf. Syst. Informatics*, vol. 3, no. 2, pp. 278–302, 2021.
- [16] Y. A. Singgalen, "Analisis Sentimen dan Pemodelan Topik dalam Optimalisasi Pemasaran Destinasi Pariwisata Prioritas



- di Indonesia,” *J. Inf. Syst. Informatics*, vol. 4, no. 1, pp. 459–470, 2021, [Online]. Available: <http://journal-isi.org/index.php/isi/article/view/171>
- [17] F. R. Mahardika, A. A. Supianto, N. Y. Setiawan, R. S. Yuwana, and E. Suryawati, “Rekomendasi Pengembangan Fasilitas Wisata Tugu Pahlawan Surabaya Melalui Visualisasi Dashboard Hasil Klasifikasi Analisis Sentimen Ulasan Pengunjung,” *J. Teknol. Inf. dan Ilmu Komput.*, vol. 9, no. 2, pp. 363–372, 2022, doi: 10.25126/jtiik.202295655.
- [18] O. Somantri and Dairoh, “Analisis Sentimen Penilaian Tempat Tujuan Wisata Kota Tegal Berbasis Text Mining,” *J. Edukasi dan Penelit. Inform.*, vol. 5, no. 2, p. 191, 2019, doi: 10.26418/jp.v5i2.32661.
- [19] L. Widiastuti, “Pemilihan Fitur Pada Analisis Sentimen Review Travel Online Menggunakan Algoritma Naïve Bayes Dalam Penerapan Mutual Information Dan Particle Swarm Optimization ( PSO ),” *IJCIT (Indonesian J. Comput. Inf. Technol.*, vol. 3, no. 1, pp. 91–100, 2018.
- [20] Murnawan and A. Sinaga, “Pemanfaatan Analisis Sentimen Untuk Pemeringkatan Popularitas Tujuan Wisata,” *J. Penelit. Pos dan Inform.*, vol. 7, no. 2, pp. 109–120, 2017, doi: 10.17933/jppi.2017.070203.
- [21] L. Wilianto, T. H. Pudjiantoro, and F. R. Umbara, “Analisis Sentimen Terhadap Tempat Wisata dari Komentar Pengunjung dengan Menggunakan Metode Naive Bayes Classifier Studi Kasus Jawa Barat,” in *Prosiding SNATIF ke-4*, 2017, pp. 439–448. doi: 10.33022/ijcs.v8i1.163.
- [22] B. Noori, “Classification of Customer Reviews Using Machine Learning Algorithms,” *Appl. Artif. Intell.*, vol. 35, no. 8, pp. 567–588, 2021, doi: 10.1080/08839514.2021.1922843.
- [23] K. M. Sukiakhy and C. V. R. Jummi, “Klasifikasi Dan Visualisasi Pariwisata Aceh Dengan Menggunakan Data Pada Twitter,” *JUSIM (Jurnal Sist. Inf. Musirawas)*, vol. 5, no. 2, pp. 148–157, 2020, doi: 10.32767/jusim.v5i02.1062.
- [24] R. Kosasih and A. Alberto, “Analisis Sentimen Produk Permainan Menggunakan Metode TF-IDF Dan Algoritma K-Nearest Neighbor,” *InfoTekJar J. Nas. Inform. dan Teknol. Jar.*, vol. 6, no. 1, pp. 134–139, 2021.
- [25] R. A. Barro, I. D. Sulvianti, and M. Afendi, “Penerapan Synthetic Minority Oversampling Technique (Smote) Terhadap Data Tidak Seimbang Pada Pembuatan Model Komposisi Jamu,” *Xplore J. Stat.*, vol. 1, no. 1, pp. 1–6, 2013.
- [26] Y. E. Kurniawati, “Class Imbalanced Learning Menggunakan Algoritma Synthetic Minority Over-sampling Technique – Nominal (SMOTE-N) pada Dataset Tuberculosis Anak,” *J. Buana Inform.*, vol. 10, no. 2, pp. 134–143, 2019, doi: 10.24002/jbi.v10i2.2441.
- [27] T. Pattiasina and D. Rosiyadi, “Comparison of Data Mining Classification Algorithm for Predicting the Performance of High School Students,” *J. Techno Nusa Mandiri*, vol. 17, no. 1, pp. 22–30, 2020, doi: 10.33480/techno.v17i1.1226.
- [28] F. Nurhuda, S. W. Sihwi, and A. Doewes, “Analisis Sentimen Masyarakat Terhadap Pilpres 2019 Berdasarkan Opini Dari Twitter Menggunakan Metode Naive Bayes Classifier,” *J. ITSMART*, vol. 2, no. 2, pp. 35–42, 2013, doi: 10.51519/journalcisa.v1i3.45.
- [29] M. F. Asshiddiqi and K. M. Lhaksana, “Perbandingan Metode Decision Tree dan Support Vector Machine untuk Analisis Sentimen pada Instagram Mengenai Kinerja PSSI,” in *e-Proceeding of Engineering*, 2020, vol. 7, no. 3, pp. 9936–9948.
- [30] R. Puspita and A. Widodo, “Perbandingan Metode KNN, Decision Tree, dan Naïve Bayes Terhadap Analisis Sentimen Pengguna Layanan BPJS,” *J. Inform. Univ. Pamulang*, vol. 5, no. 4, pp. 646–654, 2021, doi: 10.32493/informatika.v5i4.7622.
- [31] D. N. Fitriana and Y. Sibaroni, “Sentiment Analysis on KAI Twitter Post Using Multiclass Support Vector Machine (SVM),” *J. RESTI (Rekayasa Sist. dan Teknol. Informasi)*, vol. 4, no. 5, pp. 846–853, 2020, doi: 10.29207/resti.v4i5.2231.
- [32] A. Karim, “Perbandingan Prediksi Kemiskinan di Indonesia Menggunakan Support Vector Machine (SVM) dengan Regresi Linear,” *J. Sains Mat. dan Stat.*, vol. 6, no. 1, pp. 107–112, 2020, doi: 10.24014/jsms.v6i1.9259.
- [33] E. A. Nida, “Analisis Kinerja Algoritma Support Vector Machine (SVM) Guna Pengambilan Keputusan Beli/Jual Pada Saham PT Elnusa Tbk. (ELSA),” *J. Transform.*, vol. 17, no. 2, pp. 160–170, 2020, doi: 10.26623/transformatika.v17i2.1649.
- [34] Hairani, N. A. Setiawan, and T. B. Adji, “Metode Klasifikasi Data Mining dan Teknik Sampling SMOTE Menangani Class Imbalance untuk Segmentasi Customer pada Industri Perbankan,” in *Seminar Nasional Sains dan Teknologi*, 2016, pp. 168–172.
- [35] R. M. Ziku, “Partisipasi Masyarakat Desa Komodo Dalam Pengembangan Ekowisata Di Pulau Komodo,” *J. Master Pariwisata*, vol. 2, no. 1, pp. 1–21, 2015, doi: 10.24843/jumpa.2015.v02.i01.p01.
- [36] M. H. Idris, Selva, and R. Destari, “Pengaruh Destinasi Pariwisata Pulau Komodo Terhadap Beberapa Aspek Pembangunan Di Kabupaten Manggarai Barat,” *JIAP (Jurnal Ilmu Adm. Publik)*, vol. 7, no. 1, pp. 56–68, 2019, doi: 10.31764/jiap.v7i1.776.
- [37] M. N. Prakoso and A. Pratiwi, “Proses Reservasi Perjalanan Wisata Berlayar Menuju Pulau Komodo Melalui Kakabantrip Tour and Travel Service,” *J. Indones. Tour. Policy Stud.*, vol. 2, no. 2, pp. 1–21, 2017.
- [38] R. N. Chasanah and A. Wijaya, “Pengaruh Motivasi Wisata Dan Destination Image Terhadap Niat Wisatawan Untuk Berkunjung Ke ‘ 10 New Bali,’” *Value J. Manaj. dan Akunt.*, vol. 15, no. 2, pp. 268–280, 2020, doi: 10.32534/jv.v15i2.1476.